Multimodal Medical Image Synthesis and Cross-Modality Conversion for Improved Diagnosis and Explainable AI Analysis

Premanand Ghadekar   
Department of Information Technology  
Vishwakarma Institute of TechnologyPune, India  
[premanand.ghadekar@vit.edu](mailto:premanand.ghadekar@vit.edu)

Isha Kulkarni  
Department of Information Technology  
Vishwakarma Institute of TechnologyPune, India  
[kulkarni.isha23@vit.edu](mailto:kulkarni.isha23@vit.edu)Pranav Jadhav  
Department of Information Technology  
Vishwakarma Institute of TechnologyPune, India  
[pranav.jadhav23@vit.edu](mailto:pranav.jadhav23@vit.edu)

Om Lohade  
Department of Information Technology  
Vishwakarma Institute of TechnologyPune, India  
[om.lohade23@vit.edu](mailto:om.lohade23@vit.edu)Sai Kulkarni  
Department of Information Technology  
Vishwakarma Institute of TechnologyPune, India  
[sai.kulkarni23@vit.edu](mailto:sai.kulkarni23@vit.edu)

Isha Mahajan  
Department of Information Technology  
Vishwakarma Institute of TechnologyPune, India  
[isha.mahajan23@vit.edu](mailto:isha.mahajan23@vit.edu)

***Abstract*—** **Clinical diagnosis using multimodal medical images, aid in clinical diagnosis of disease status. Medical image synthesis alleviates this by creating artificial images that resemble those acquired through conventional imaging methods. In this work, we aim to synthesize chest X-ray images and generate brain CT from MRI data by proposing a novel multimodal image synthesis approach with the use of GANs. This is achieved by integrating visual data with text reports. The proposed model can also be used for generating synthetic images for other diseases. The framework achieves better training stability and inherent structural consistency by resolving issues like model collapse and instability of gradients. The chest x-ray synthesis model and the MRI to CT model achieved an SSIM score of 0.71 and 0.75 respectively. An FID score of 20, Inception score of 8.5, CLIP similarity of 0.95, Grad CAM interpretability of 0.90, Perceptual similarity of 0.92 and a training stability of 0.87 was achieved by the image synthesis model indicating that it performed better than traditional GANs.**

***Keywords–*** *MRI, CT, X-rays, Multimodal, Image Synthesis*

# INTRODUCTION

AI in healthcare imaging helps detect and predict diseases. CT, MRI, and X-ray are primary imaging techniques, but large datasets required for AI model development often remain insufficient. Multimodal medical image synthesis addresses this issue by generating synthetic images to supplement real datasets, improving AI-based diagnostics.

Radiological imaging aids disease diagnosis and treatment planning. While CT provides better bone detail and MRI offers superior soft tissue contrast, challenges such as cost, availability, and radiation risks persist. To overcome these limitations, medical image synthesis techniques, including deep learning-based methods, have been developed [1].

Multimodal medical imaging integrates MRI, CT, PET, and X-rays, offering a comprehensive view of diseases. This fusion enables precise diagnosis, optimized treatment plans, and improved patient outcomes by leveraging multiple imaging sources [2].

Our study focuses on multimodal image synthesis, particularly chest X-ray generation and MRI-to-CT conversion. Synthetic X-rays improve AI model training, while MRI-to-CT conversion is useful in cases where CT scans are inaccessible, producing high-contrast CT-like images for better diagnosis.

Obtaining multiple contrast MRI scans is costly and time-consuming. Synthetic imaging provides an alternative by generating artificial contrasts, reducing scanning requirements and enhancing imaging efficiency [3].

Recent advancements in medical imaging incorporate multimodal foundation models that integrate text, images, and patient records for improved diagnostics. While these models enhance performance, research highlights barriers to widespread adoption that need further refinement [4].

Explainable AI (XAI) is becoming essential in healthcare, ensuring model transparency and reliability. Studies emphasize its role in enhancing interpretability, building user trust, and improving AI accountability in clinical applications [5].

This research introduces a multimodal image synthesis framework focusing on MRI-to-CT transformation and synthetic chest X-ray generation. Additionally, it integrates XAI for validation, ensuring interpretability for clinicians and enhancing the reliability of AI-driven diagnostic models.

II.LITERATURE REVIEW

Medical image synthesis helps address the challenge of acquiring multiple imaging modalities for clinical evaluations by generating PET, CT, and MRI images from existing data. Recent deep learning advancements have outperformed traditional methods in artificial image contrast generation. A survey from 2018 to 2023 discusses model architectures, loss functions, datasets, and performance metrics while outlining challenges and future directions [6].

Techniques for estimating attenuation-coefficient maps from MR images have improved PET-MR systems and MR-guided radiotherapy. Traditional segmentation and atlas-based corrections struggle with tissue differentiation and anatomical accuracy. CNNs show promise but sometimes fail to preserve structure, while VAEs require better fidelity. GANs with modality-invariant constraints enhance cross-modality image synthesis [7].

Multimodal image synthesis and editing use deep learning techniques like conditional GANs and diffusion models to model interactions between modalities. Despite advancements, high-resolution synthesis, feature alignment, and evaluation metrics remain challenging. Improved feature fusion and evaluation frameworks are needed [8].

Medfusion, a latent conditional denoising diffusion probabilistic model (DDPM), was tested on radiology, histopathology, and ophthalmology datasets, outperforming state-of-the-art GANs in fidelity and diversity [9].

Multimodal image synthesis methods are categorized into diffusion-based, autoregressive, and GAN-based models. While models like CLIP, DALL-E, and DALL-E-2 show potential, they face challenges such as bias, misinformation risks, and coherence issues [10].

Deep learning has enhanced organ and lesion identification in medical imaging, improving diagnostic accuracy. However, the black-box nature of AI models limits adoption in clinical workflows. Explainable AI (XAI) techniques, including concept-based and visual explanation methods, help improve model transparency [11].

A survey on XAI in medical diagnosis and surgery highlights techniques like LIME, SHAP, and class activation mapping for disorders such as breast cancer and Alzheimer's. XAI also plays a role in surgical skill evaluation and laser surgery selection, reinforcing AI interpretability [12].

Multimodal imaging techniques in cardiovascular disease integrate MRI, CT, and echocardiography using advanced segmentation and fusion models. Studies demonstrate improved diagnostic accuracy but highlight challenges in integrating additional modalities, such as X-ray and non-imaging data [13].

To overcome limitations of conditional image synthesis models, the Product-of-Experts GAN (PoE-GAN) was introduced. This model synthesizes high-quality images using multiple input modalities, such as text, segmentation, and sketches, improving image diversity and quality [14].

Multimodal Image Synthesis and Editing (MISE) research explores deep neural networks addressing the "modality gap" between different data types like text, audio, and images. Improving feature association enhances AI-generated creativity, making it more intuitive [15].

A novel multimodal image synthesis framework, MMoT, tackles coordination issues in image construction. Through extensive testing, it demonstrated superior performance in generating high-quality images, even when handling mismatched input modalities [16].

Multimedia image synthesis integrates text descriptions, different image formats, and sound to enhance medical imaging and content creation. Models like VAEs and GANs improve image accuracy by synthesizing diverse input patterns, benefiting diagnostic applications [17].

Studies on integrating MRI, CT, and ultrasound images propose computational methods that help physicians diagnose conditions more effectively. These methods lay the foundation for further AI-based advancements in medical imaging [18].

Deep learning techniques combining textual and image data generate realistic medical images, addressing challenges like limited high-quality datasets for AI training. Research highlights key challenges and future trends in synthetic medical imaging [19].

DCGANs have been optimized for medical image generation, allowing models to learn from real medical images and create enhanced versions. This approach improves data availability for AI training, benefiting diagnostic applications in medical science and education [20].

A study on synthetic data augmentation emphasizes how technology can further enhance medical imaging and healthcare solutions. The research highlights DCGANs' practical applications in refining AI diagnostic tools and expanding dataset availability for clinical applications [21].

# METHODOLOGY

1. ***Multimodal Synthesis of Chest X-Ray Images***



1. *Dataset Description*

The dataset used in this study consists of approximately 7,000 chest X-ray images paired with corresponding XML reports. This dataset was selected for its rich multimodal content, providing both visual and textual information essential for generating customized synthetic images.

Each X-ray image is paired with a corresponding XML report generated by radiologists. These reports contain detailed findings, impressions, and diagnostic conclusions that describe the observed pathologies and anatomical observations. The XML structure follows a consistent schema, with specific fields for:

* **Findings**: Descriptive information on the anatomical regions affected, including specific observations related to lung fields, pleura, and mediastinal structures.
* **Impression**: A concise summary of the radiologist’s interpretation, indicating any suspected or confirmed diagnoses.
* **Additional Notes**: Optional fields for further clinical observations, recommendations, and follow-up considerations.

Text mining techniques were applied to extract relevant medical terms and conditions from each XML report. Key indicators of diseases (e.g., pneumonia, pleural effusion, atelectasis) were identified and standardized to ensure consistency across the dataset.

1. *Model Architecture:*

For multimodal synthesis, a modified GAN (Generative Adversarial Network) architecture was employed, utilizing two primary input branches:

* **Image Encoder**: The image input branch processes chest X-rays through a convolutional neural network (CNN) encoder, extracting visual features that represent spatial and structural characteristics of the lungs and surrounding anatomy.
* **Text Encoder**: The text input branch utilizes a transformer-based encoder to parse and process the XML-formatted medical reports. This encoder generates a semantic vector that represents the radiological descriptions, allowing the synthesis model to condition the generated output on these features.

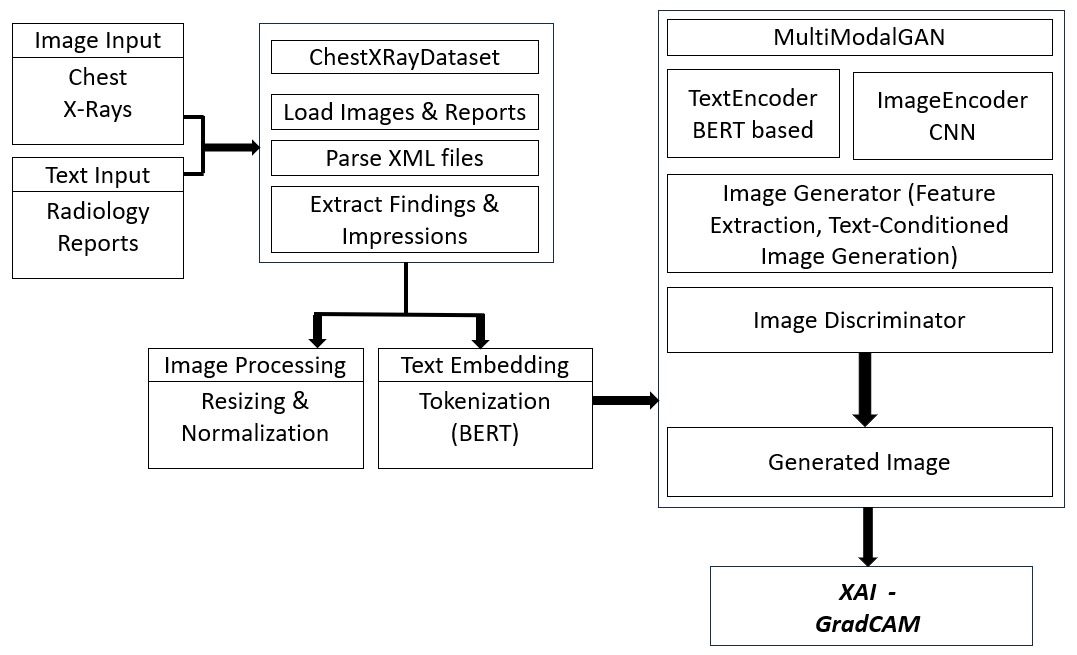


Fig 1. Working of proposed MultiModalGAN framework

1. *Fusion Mechanism:*

The fusion mechanism combines visual and textual embeddings as shown in Fig 1. to create a unified multimodal representation that conditions the generation of synthetic X-ray images. Let ​ denote the visual embedding from the image encoder and ​ ​ the textual embedding from the text encoder. These embeddings are combined in a fusion module:

(1)

where ​ represents a fully connected layer followed by a non-linear activation (ReLU). The fused embedding ​ captures both spatial characteristics from the X-ray image and semantic information from the textual report, providing a rich feature representation that guides the generation process.

1. *Image generation process:*

The image generation process begins with a MultiModalGAN model, which synthesizes X-ray images conditioned on both visual and textual inputs. As shown in Fig 2., the generator in this model takes as input a combination of features derived from an existing X-ray image and a text description, represented by the embeddings extracted from textual data (radiology reports). This multimodal input enables the generator to create a synthetic X-ray image that aligns with specific anatomical and pathological details provided by the user.

A training scheme is of adversarial nature, where a discriminator and generator are optimized to make generated images more realistic. This leads to the following steps as to how the loss functions contribute:

* Adversarial loss: The adversarial framework is, therefore, used to enforce that the fake X-ray images created by the technique are similar to the real body X-ray images. The optimization function called Binary Cross-Entropy Loss, which is the key to helping the generator to produce images, which the discriminator cannot determine are fake or real X-rays. The decision-making algorithm will be the one that we train to learn to recognize the difference between the set of real and simulated images.In case of real images, the output of the discriminator is optimized to obtain the value close to 1 (truth label), and for synthetic images, the output is optimized to be close to 0 (fake label). The discriminator loss is computed as:

(2)

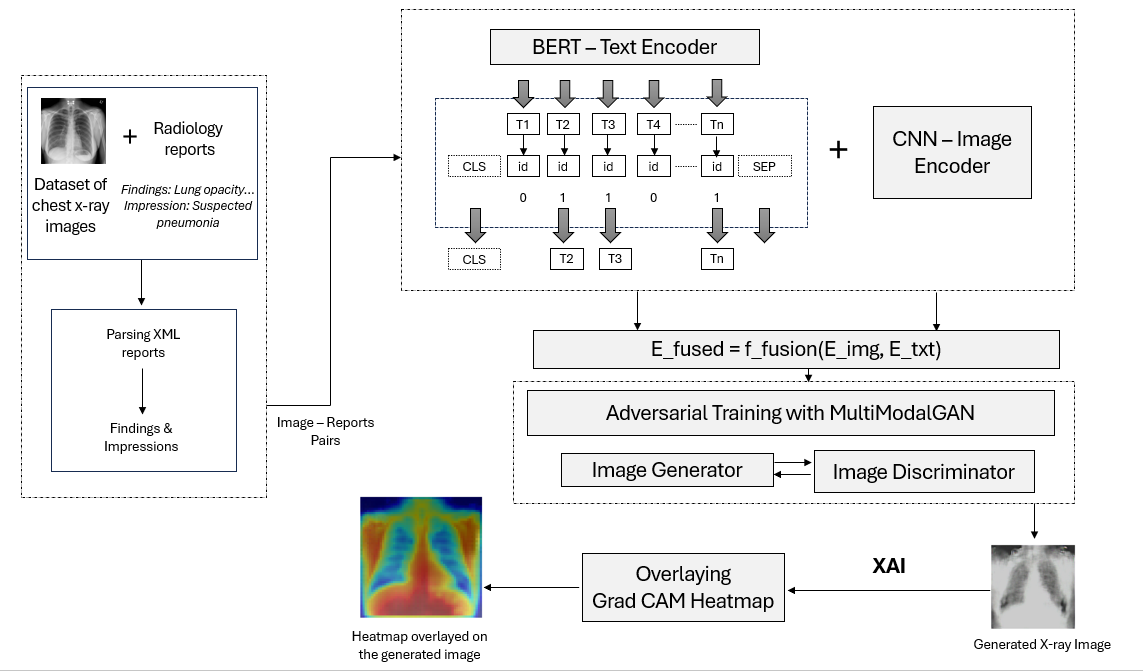


Fig. 2. Overview of MultimodalGAN architecture using BERT-based and CNN-based encoders



where D represents the discriminator, is the real image, G is the generator, and ​ is the multimodal embedding input.

The generator just deals with the production of images. However, it’s trained by the discriminator to make the generated images look as realistic as possible. The adversarial loss component of the generator is:

(3)

To make synthetic images not only realistic but also contain actual anatomical characteristics from the input, a Mean Squared Error (MSE) Loss is implemented. This part of the loss lowers pixel-to-pixel differences between the fake and real images, thereby making the model stick to the anatomical structure. The content similarity loss is defined as:

(4)

The total loss for the generator combines the adversarial and content similarity losses, weighted to balance realism and anatomical accuracy. This is expressed as:

(5)

where λ is a weighting factor set to control the influence of content similarity, promoting image fidelity to real X-rays. A residual block is added to the image generation process which helps prevent the vanishing gradients during the back propagation during the image generation process. Additionally, by optimizing these loss functions, the generator learns to create synthetic X-ray images that are both visually realistic and structurally aligned with the provided descriptions, effectively synthesizing images for disease prediction and analysis.

*D. Analysis using XAI:*

We utilized, in this research project, explainable AI (XAI) techniques to deconstruct and confirm the produced chest X-ray images which enabled us to make sure the outputs of the model correspond to the clinically relevant features. We have used the CAM and the Gradient-weighted CAM (Grad-CAM) to create images that showed the regions that the model mainly used for its predictions (sect. 3D). Such methods allowed us to enquiry whether generated images were specific pathological markers or segments like lung filtering which, therefore, can then be associated with disorders such as pneumonia or pleural effusion.

To further understand the model's internal processes, we utilized hooks to capture intermediate feature maps, revealing which layers focused on key anatomical structures. Lastly, we overlaid the generated heatmaps onto the original or generated images, creating composite visualizations that allow for easy comparison of the highlighted areas with actual anatomical regions. This XAI approach ensures that the model produces realistic and diagnostically relevant images, offering a transparent and interpretable synthesis process critical for medical applications.

The complete image synthesis process is shown below:

|  |
| --- |
| **Algorithm for MultiModal Image Synthesis** |
| ***Step 1***: Parse the XML radiology reports to extract Findings and Impressions.  ***Step 2****:* Extract visual features and text embeddings in the form of tokens and combine them:   1. *BERT-based encoder:*    1. Tokenize the text and add special tokens CLS and SEP to the start and end respectively.    2. Assign the importance score (0 or 1) after analysing the semantic relationship between the tokens. 2. *CNN-based image encoder:*   Extract the spatial features from the input image using convolutional layers and encode the image for further processing.  ***Step 3****:* Pass the fused features to the conditional GAN model.  ***Step 4:*** Add a residual block. Pass the fused features through series of layers in the network and use skip connection for enhanced feature flow and propagation.    ***Step 5****:* Train the model using the fused input data and use it for synthesizing images.  ***Step 6:*** Generate the heatmap using Grad CAM. Overlay it on the generated image to identify and analyse the parts of the image that contributed the most to the image synthesis process. |

Top of Form

Bottom of Form

1. ***Brain MRI and CT cross-modality conversion***

The second part of this research focuses on translating brain MRI images to CT scans using CycleGAN, an unsupervised deep learning approach capable of translating between two unpaired domains, such as MRI and CT modalities, without paired examples in the training set. CycleGAN is particularly effective for this conversion, as it preserves anatomical consistency while adapting the input modality to resemble the target modality.

* 1. *Dataset Description*

This MRI-to-CT dataset contains approximately 1,742 MRI and CT image pairs, each representing the same anatomical regions. These images provide cross-modality pairs that guide the CycleGAN model in learning mappings from MRI to CT scans, enhancing the quality and realism of generated CT-like images.

* 1. *Components of architecture*

The CycleGAN model works by training two generators and two discriminators where each generator tries to create pictures that are identical to real ones.

**Generators**: We have two generators:

* G: MRI→CT: Converts MRI images to CT-like images.
* F: CT→MRI: Converts CT images back to MRI-like images.

**Discriminators**: The quality of the produced images is assessed by two discriminators:

* DCT​: Checks if the generated CT images look like real CTs.
* DMRI​: Checks if the generated MRI images look like real MRIs.

The training process uses the Cycle Consistency Loss function. This loss ensures that if we convert an MRI to a CT image and then back to an MRI, the resulting MRI should closely resemble the original. This helps the model keep anatomical consistency across translations.

(6)

Here, F(G(MRI)) should match the original MRI, and G(F(CT)) should match the original CT. The ∥⋅∥ term (L1 norm) measures pixel-wise similarity.

|  |
| --- |
| **Algorithm for MRI and CT cross-modality conversion** |
| ***Step 1***: Define 2 generators G, F and 2 discriminators DCT and DMRI.  ***Step 2:*** Setup the adversarial loss and cycle consistency loss functions.  ***Step 3:*** Use G to generate CT images from MRI and use F to generate MRI images from CT.  G: MRI→CT  F: CT→MRI  ***Step 4****:* Rebuild MRI and CT following the cycle consistency step.  ***Step 5:*** Begin DCT training and DMRI to differentiate between real CT, and real MRI and .  ***Step 6:*** Use the cycle consistency loss and adversarial loss to optimize G and F. Repeat the steps 2 to 5 until losses converge. |

# RESULTS AND DISCUSSION

* 1. Multimodal Chest X-Ray Synthesis

The multimodal GAN model for chest X-ray synthesis generated synthetic images conditioned on specific textual inputs from radiology reports. Quantitatively, the generated images achieved a structural similarity index (SSIM) of over 0.71 when compared with real images, indicating a strong alignment in visual quality.

The integration of text-based conditioning proved effective, with Grad-CAM and CAM heatmaps showing that the model focused on appropriate anatomical regions based on the descriptions. This multimodal approach highlights the model’s potential to generate synthetic data with clinically relevant characteristics, which could be useful for training other AI models. Fig 3. shows the inputs and their corresponding generated images.

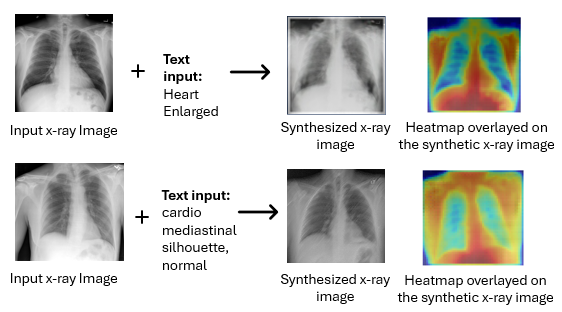


Fig. 3. Chest x-ray images generated using MultiModalGAN

Table 1: Comparison based on SSIM metric

|  |  |  |
| --- | --- | --- |
| Quality Metric | MultiModalGAN | CycleGAN |
| SSIM | 0.71 | 0.75 |

The Structural Similarity Index Measure (SSIM) calculated for both chest x-ray synthesis and brain MRI to CT conversion is shown in Table 1. Fig 4. shows the generator and discriminator loss over the epochs showing model’s stability.

Explainable AI provided critical observations in model's decision-making process. Using CAM and Grad-CAM, we confirmed that the generated images emphasized relevant anatomical regions, with heatmaps overlaying correctly on areas associated with the conditions mentioned in the textual inputs. These visualizations verified the clinical relevance of the synthetic images, adding a layer of transparency to the model’s outputs. This XAI analysis not only builds trust in the model’s accuracy but also reinforces the potential of using generated images in diagnostic workflows.

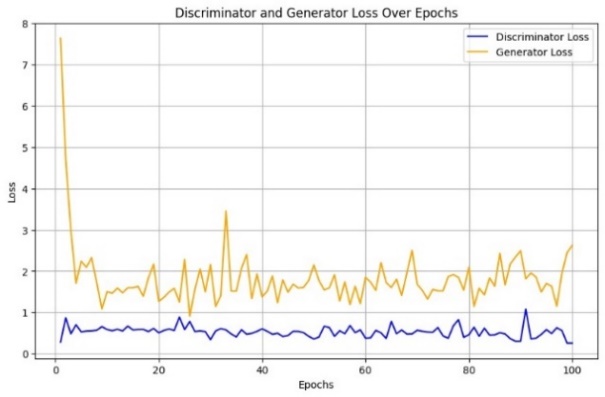


Fig. 4. Discriminator and Generator Loss over epochs

* 1. MRI to CT conversion

The MRI-to-CT conversion model, built on CycleGAN, achieved visually convincing CT-like images from MRI inputs, with minimal structural loss. Visual inspection of the generated CT images demonstrated that the model preserved essential details while adapting the modality. The SSIM score achieved was around 0.75. These results indicate that the model can produce CT-like images as shown in Fig 5. with a high degree of anatomical accuracy, making it a valuable tool in scenarios where CT scans are limited or unavailable.

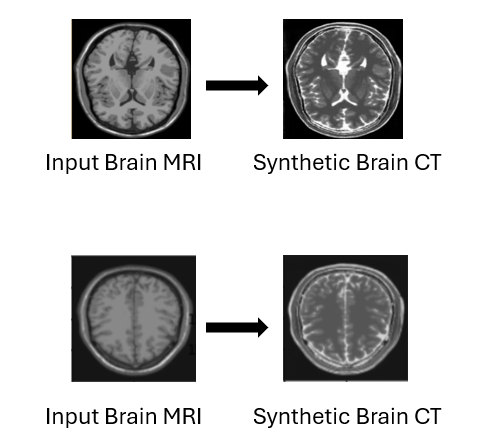


Fig. 5. Brain MRI-CT conversion using CycleGAN

The results are compared with some existing models and the results of the comparison are shown in Table 2.

Table 2. Performance Comparisons with Existing Models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | FID Score | Inception Score | CLIP Similarity | Grad-  CAM Interpretability | Perceptual Similarity | Training Stability |
| Multi Modal  GAN | 20 | 8.5 | 0.95 | 0.90 | 0.92 | 0.87 |
| StyleGAN | 40 | 7.8 | 0.88 | 0.85 | 0.86 | 0.80 |
| WGAN-GP | 35 | 7.5 | 0.90 | 0.86 | 0.89 | 0.83 |

# CONCLUSION

This study successfully demonstrates the use of multimodal and cross-modality synthesis for medical imaging, focusing on two main objectives: generating synthetic chest X-rays conditioned on textual descriptions and transforming MRI images into CT-like representations. By employing a   
multimodal GAN model, the generated chest X-rays align well with both the anatomical and pathological descriptions provided in radiology reports.

Additionally, the CycleGAN-based MRI-to-CT conversion model enables an efficient transformation between modalities, facilitating applications where CT images may be needed but are not available. Together, these models enhance data accessibility, expand opportunities for synthetic data generation, and hold potential to support diagnostic practices in scenarios with limited imaging data.

Explainable AI techniques, including CAM, Grad-CAM, and overlay visualizations, have been crucial for validating the relevance of generated images, ensuring interpretability and adherence to clinical expectations

REFERENCES

1. *HLSNC-GAN: Medical Image Synthesis Using Hinge Loss and Switchable Normalization in CycleGAN*. (2024). ieee.org. https://ieeexplore.ieee.org/document/10504113/
2. mostafa.elhabibdahouniv-brest.fr. (2024). A review of deep learning-based information fusion techniques for multimodal medical image classification. arxiv.org. https://arxiv.org/html/2404.15022v1
3. Dai, X., Lei, Y., Fu, Y., Curran, W. J., Liu, T., Mao, H., & Yang, X. (2020). Multimodal MRI synthesis using unified generative adversarial networks. *Medical Physics*. <https://doi.org/10.1002/mp.14539>
4. Huang, S.-C., Jensen, M., Yeung-Levy, S., Lungren, M. P., Poon, H., & Chaudhari, A. S. (2024). Multimodal Foundation Models for Medical Imaging - A Systematic Review and Implementation Guidelines. https://doi.org/10.1101/2024.10.23.24316003
5. Yang, W., Wei, Y., Wei, H., Chen, Y., Huang, G., Li, X., Li, R., Yao, N., Wang, X., Gu, X., Amin, M. B., & Kang, B. (2023). Survey on Explainable AI: From Approaches, Limitations and Applications Aspects. Human-Centric Intelligent Systems. <https://doi.org/10.1007/s44230-023-00038-y>
6. Sanuwani Dayarathna, Kh Tohidul Islam, Sergio Uribe, Guang Yang, Munawar Hayat, Zhaolin Chen,Deep learning based synthesis of MRI, CT and PET: Review and analysis,Medical Image Analysis,Volume 92,2024,103046,ISSN 1361-8415,https://doi.org/10.1016/j.media.2023.103046.
7. Reaungamornrat S, Sari H, Catana C, Kamen A. Multimodal image synthesis based on disentanglement representations of anatomical and modality specific features, learned using uncooperative relativistic GAN. Med Image Anal. 2022 Aug;80:102514. doi: 10.1016/j.media.2022.102514. Epub 2022 Jun 11. PMID: 35717874; PMCID: PMC9810205.
8. F. Zhan et al., "Multimodal Image Synthesis and Editing: The Generative AI Era," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 12, pp. 15098-15119, Dec. 2023, doi: 10.1109/TPAMI.2023.3305243.
9. Müller-Franzes, G., Niehues, J. M., Khader, F., Arasteh, S. T., Haarburger, C., Kuhl, C., Wang, T., Han, T., Nolte, T., Nebelung, S., Kather, J. N., & Truhn, D. (2023). A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis. Scientific Reports. https://doi.org/10.1038/s41598-023-39278-0
10. Shahrooz Faghihroohi. (2024). Multimodal Image Synthesis - DLMA: Deep Learning for Medical Applications - BayernCollab. dvb.bayern. https://collab.dvb.bayern/display/TUMdlma/Multimodal+Image+Synthesis
11. Pahud de Mortanges, A., Luo, H., Shu, S. Z., Kamath, A., Suter, Y., Shelan, M., Pöllinger, A., & Reyes, M. (2024). Orchestrating explainable artificial intelligence for multimodal and longitudinal data in medical imaging. Npj Digital Medicine. https://doi.org/10.1038/s41746-024-01190-w
12. cristiano.patricioubi.pt. (2023). Explainable Deep Learning Methods in Medical Image Classification: A Survey. arxiv.org. https://arxiv.org/html/2205.04766
13. Zhang, Y., Weng, Y., & Lund, J. (2022). Applications of Explainable Artificial Intelligence in Diagnosis and Surgery. Diagnostics. https://doi.org/10.3390/diagnostics12020237
14. Milosevic, M., Jin, Q., Singh, A., & Amal, S. (2024). Applications of AI in multi-modal imaging for cardiovascular disease. Frontiers in Radiology. https://doi.org/10.3389/fradi.2023.1294068
15. Multimodal Conditional Image Synthesis with Product-of-Experts GANs. (2023). nvidia.com. https://research.nvidia.com/labs/dir/PoE-GAN/
16. Multimodal Image Synthesis and Editing: The Generative AI Era. (2024). ieee.org. <https://ieeexplore.ieee.org/document/10230895/>
17. Jianbin Zheng. (2023). MMoT: Mixture-of-Modality-Tokens Transformer for Composed Multimodal Conditional Image Synthesis. arXiv.org. https://arxiv.org/abs/2305.05992v1
18. fnzhan. (2024). GitHub - fnzhan/Generative-AI: [TPAMI 2023] Multimodal Image Synthesis and Editing: The Generative AI Era. GitHub. https://github.com/fnzhan/Generative-AI
19. Englmeier, K.-H., Haubner, M., Fink, U., & Fink, B. (1994). Image analysis and synthesis of multimodal images in medicine. Computer Methods and Programs in Biomedicine. https://doi.org/10.1016/0169-2607(94)90070-1
20. luosanbiiie.ac.cn. (2024). A Survey on Multimodal Deep Learning for Image Synthesis. acm.org. https://dl.acm.org/doi/fullHtml/10.1145/3461353.3461388
21. Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., & Shen, D. (2018). Medical Image Synthesis with Deep Convolutional Adversarial Networks. IEEE Transactions on Biomedical Engineering. https://doi.org/10.1109/tbme.2018.2814538